

УДК 534.4:621.391

О.М. Карпов, К.М. Глушак

Дніпропетровський національний університет імені Олеся Гончара

ВИЗНАЧЕННЯ ЗНАЧУЩИХ ПАРАМЕТРІВ МОВНОГО СИГНАЛУ ЗА ДОПОМОГОЮ КОЛОКОЛОПОДІБНИХ ФУНКЦІЙ

Наведений один із можливих варіантів технології визначення функціонального стану людини за його мовою на підставі отримання різниці спектрально-часових представлень, а також їх відношень.

Ключові слова: *функціональний стан, людина, мова, настрій, розподіл частот, значущі параметри, апроксимація, колоколоподібні функції, функція Аньєзі.*

Приведен один из возможных вариантов технологии определения функционального состояния человека по его речи на основании получения разности спектрально-временных представлений, а также их отношений.

Ключевые слова: *функциональное состояние, человек, речь, настроение, распределение частот, аппроксимация, колоколообразные функции, функция Аньези.*

This work tells us how to determine the functional state of a person according to his speech. Which is based on the finding difference between two wavelength-time representations.

Keywords: functional state, human, speech, mood, frequency distribution, witch of agnesy, approximation.

Постановка задачі. Авторами був виконаний ряд досліджень при реалізації пристроїв розпізнавання промови.

Задача побудови пристроїв розпізнавання для сучасних персональних ЕОМ складається із двох частин:

1) вибір параметрів опису об'єктів (фонем, складів, слів, фраз), а також методів інтерпретації описів і проектування програмних систем, побудови програмних засобів реалізації описів і розпізнавання;

2) вбудовування розроблених проектів і програмних реалізацій у системні середовища (оболонки) ПЕОМ.

Останніми роками рівень розвитку обчислювальних засобів значно виріс, що дозволило вирішувати задачі побудови систем розпізнавання із залученням великого числа параметрів і методів, орієнтованих на глибинну сутність мовного сигналу.

Однак залишилися необхідність узгодження (вирішення) суперечливих вимог: забезпечення високої надійності розпізнавання для великих словників об'єктів, що потребує залучення великого числа параметрів (ознак) і, відповідно, великого часу їхньої обробки; виконання обробки в реальному масштабі часу (мінімальні витрати часу).

У цілому задача проектування пристрою розпізнавання мовлення може бути сформульована в такий спосіб:

а) вибрати таку систему параметрів, що забезпечувала б досить повний опис деякої заданої кількості об'єктів мовних сигналів та їхню відмінність один від одного в цій системі з деякою заданою надійністю; параметри повинні обчислюватися зі швидкістю, що має реальний час функціонування пристрою розпізнавання мовлення; сукупність параметрів кожного об'єкта повинна займати такий обсяг пам'яті, щоб забезпечувати можливість розміщення в ній еталонних параметрів досить великих словників (наприклад, одночасне розміщення не менш 500–1000 слів із параметрами);

б) здійснити сегментацію мовлення і групування інформації про параметри для формування структури слова і завдання відповідності між символьним і параметричним представленнями;

в) побудувати стратегію розпізнавання, що забезпечує мінімальні витрати часу на пошук найбільш відповідного еталона з максимальною надійністю в заданій системі параметрів;

г) розробити й реалізувати проект пристрою розпізнавання мовлення, що має можливість подальшого розвитку (розширення або звуження) із мінімальними витратами на його модифікацію.

У результаті теоретичних та експериментальних досліджень була розроблена загальна схема проектування пристроїв розпізнавання мовлення [1; 2], що містить складові частини:

- блок аналізу мовних сигналів;
- вибір типів мовних одиниць (сегменти, склади, фонemi, слова, сегментно-складове представлення);
- сегментація;
- вибір метрики та критеріїв порівняння;
- вибір методів порівняння.

При проектуванні складних програмних систем необхідно вирішити такі практичні задачі:

а) вибір універсального мовного середовища для написання модулів будь-якої складності і будь-якого функціонального призначення;

б) забезпечення мовних засобів для маніпулювання текстами до їхньої трансляції;

в) розробка технології проектування програмних систем, що містять параметри їхньої перебудови по вихідних текстах програм.

Алгоритм апроксимації на основі колоколоподібних функцій спектрально-часового представлення мовних сигналів

Задача апроксимації спектрально-часового представлення в класі колоколоподібних функцій полягає в розбивці заданого початкового опису на ділянки, що відповідають деяким мовним одиницям, і наступний їхній опис у класі заданих функцій. Вихідне спектрально-часове представлення розглядається у прямокутній області $R = [\omega_a, \omega_b] \times [T_c, T_d]$, в області R задана таблична спектрально-часова функція $S(\omega_k, t_l)$, де ω_k – дискретно задана частота, t_l – дискретно заданий час: $\omega_a = \omega_0 < \omega_1 < \dots < \omega_m = \omega_b$,

$$T_c = t_0 < t_1 < \dots < t_n = T_d.$$

Визначено класи функцій $\{W_i(\omega_k)\}$, $\{h_i(t_l)\}$, що мають такі властивості:

- для представлення резонансу функція повинна бути колоколоподібною (мати максимум на заданій частоті);
- форма функції повинна визначатися деякими параметрами;
- функції асимптотично наближаються до площини області R у будь-якому напрямку від максимуму;
- гілки функції повинні монотонно убувати (зростати), якщо максимум колоколу знаходиться за межами області R.

Функція $S(\omega_k, t_l)$ в області R містить довільну кількість сплесків спектральної енергії, розміщених довільним чином у заданій області. Необхідно функцію $S(\omega_k, t_l)$ сегментувати і апроксимувати у класі функцій $\{W_i(\omega_k)\}$, для цього визначити параметри сплесків функції

$S(\omega_k, t_l)$ як параметри функцій $\{h_i, (t_i)\}$. Для рішення задачі в області R будується сітка $M \times N$, де M – кількість сегментів, N – кількість смуг по частоті. Клітки сітки визначають початкові значення параметрів по частоті та часові.

Спектрально-часова складова шукається у класі функцій модифікованого локона Аньєзі

$$(\omega^2 + r^2)W(\omega) - a^3 = 0, \quad W(\omega_k) = \frac{a_{(i)}^3}{c_{(i)}^2 + (\omega_k - \beta_{(i)})^2},$$

$$h(t_l) = \frac{b_{(j)}^3}{d_{(j)}^2 + (t_l - T_{(j)})^2},$$

де r, c, d визначають форму локона у вигляді

$$\begin{aligned} Z_{(i,j)}(k, q_{(i,j)}, a_{(i,j)}, c_{(i,j)}, \beta_{(i,j)}, b_{(i,j)}, d_{(i,j)}, T_{(i,j)}, \omega_k, t_l) = \\ = S_{g(j)}(\omega_k) W_{(i)}(\omega_k) h_{(j)}(t_l) = \\ = \frac{q_{(i)}^{\ln k}}{k} \frac{a_{(i)}^3}{c_{(i)}^2 + (\omega_k - \beta_{(i)})^2} \frac{b_{(j)}^3}{d_{(j)}^2 + (t_l - T_{(j)})^2}. \end{aligned}$$

Різницева схема вилучення складових – СЕТ [2] для спектральної функції $S(\omega_k, t_l)$:

$$\begin{aligned} S(\omega_k, t_l) = \\ = Z_{(1,1)}(q_{(1,1)}, a_{(1,1)}, c_{(1,1)}, \beta_{(1,1)}, b_{(1,1)}, d_{(1,1)}, T_{(1,1)}, \omega_k, t_l) + S_{1,1}(\omega_k, t_l), \\ S_{1,1}(\omega_k, t_l) = \\ = S(\omega_k, t_l) - Z_{(1,1)}(q_{(1,1)}, a_{(1,1)}, c_{(1,1)}, \beta_{(1,1)}, b_{(1,1)}, d_{(1,1)}, T_{(1,1)}, \omega_k, t_l) = \\ = Z_{(1,2)}(q_{(1,2)}, a_{(1,2)}, c_{(1,2)}, \beta_{(1,2)}, b_{(1,2)}, d_{(1,2)}, T_{(1,2)}, \omega_k, t_l) + S_{1,2}(\omega_k, t_l), \\ \dots \dots \dots \\ S_{i-1,j-1}(\omega_k, t_l) = \\ = Z_{(i,j)}(k, q_{(i,j)}, a_{(i,j)}, c_{(i,j)}, \beta_{(i,j)}, b_{(i,j)}, d_{(i,j)}, T_{(i,j)}, \omega_k, t_l) + S_{i,j}(\omega_k, t_l). \quad (1) \end{aligned}$$

Вилучення складових $S_{i,j}(\omega_k, t_l)$ здійснюється для кожної клітки сітки $M \times N$, $j = \overline{0, M-1}$, $i = \overline{0, N-1}$. За методом найменших квадратів визначаються параметри $k, q_{(i,j)}, a_{(i,j)}, c_{(i,j)}, \beta_{(i,j)}, b_{(i,j)}, d_{(i,j)}, T_{(i,j)}$ для

$$\sigma_{(i,j)}^2 = \sum_{\omega_k} \sum_{t_l} [S_{i-1,j-1}(\omega_k, t_l) -$$

$$-Z_{(i,j)}(k, q_{(i,j)}, a_{(i,j)}, c_{(i,j)}, \beta_{(i,j)}, b_{(i,j)}, d_{(i,j)}, T_{(i,j)}, \omega_k, t_l) \Gamma^2.$$

На першому кроці параметри визначаються для

$$\begin{aligned} q_{(1,1)}^2 &= \sum_{\omega_k} \sum_{t_l} [S(\omega_k, t_l) - \\ &- Z_{(1,1)}(q_{(1,1)}, a_{(1,1)}, c_{(1,1)}, \beta_{(1,1)}, b_{(1,1)}, d_{(1,1)}, T_{(1,1)}, \omega_k, t_l)]^2 = \\ &= \sum_{\omega_k} \sum_{t_l} \left\{ S(\omega_k, t_l) - \frac{q_{(1)}^{\ln k}}{k} \frac{a_{(1)}^3}{c_{(1)}^2 + (\omega_k - \beta_{(1)})^2} \frac{b_{(1)}^3}{d_{(1)}^2 + (t_l - T_{(1)})^2} \right\}^2, \\ \sigma &= S(\omega_k, t_l) - \frac{q_{(1)}^{\ln k}}{k} \frac{a_{(1)}^3}{c_{(1)}^2 + (\omega_k - \beta_{(1)})^2} \frac{b_{(1)}^3}{d_{(1)}^2 + (t_l - T_{(1)})^2}. \end{aligned} \quad (2)$$

Параметри $q_{(i,j)}, a_{(i,j)}, c_{(i,j)}, \beta_{(i,j)}, b_{(i,j)}, d_{(i,j)}, T_{(i,j)}$ визначаються на наступних кроках.

Для рішення задачі мінімізації (2) реалізований метод покоординатного спуску, що полягає у виборі напрямку зміни параметрів $k, q_{(i,j)}, a_{(i,j)}, c_{(i,j)}, \beta_{(i,j)}, b_{(i,j)}, d_{(i,j)}, T_{(i,j)}$ за правилом

$$\begin{aligned} q_{(i,j)} &= q_{0(i,j)} \pm \Delta_q, \quad a_{(i,j)} = a_{0(i,j)} \pm \Delta_a, \quad c_{(i,j)} = c_{0(i,j)} \pm \Delta_c, \quad \beta_{(i,j)} = \beta_{0(i,j)} \pm \Delta_\beta, \\ b_{(i,j)} &= b_{0(i,j)} \pm \Delta_b, \quad d_{(i,j)} = d_{0(i,j)} \pm \Delta_d, \quad T_{(i,j)} = T_{0(i,j)} \pm \Delta_T \end{aligned}$$

із заданою точністю EPS.

Для кожного виду колоколоподібних функцій існує свій параметр, що керує шириною відповідної функції щодо положення її максимуму. Цей параметр є аналогом добротності резонансних ланок. Для локона Аньєзі такими параметрами є: $c_{(i,j)}$ – за частотою, $d_{(i,j)}$ – за часом.

Базовий алгоритм розпізнавання на основі екстремальних функцій

У результаті рішення задачі сегментації формується деяка тимчасова послідовність, що відображає структуру мовного висловлення. В основу алгоритму розпізнавання покладене рішення задачі синтезу спектрально-часового представлення в базисі аналітичних функцій двовимірного локона Аньєзі, Гауса та ін. Для цього в частотній області діапазон частот від 100 Гц до 6 КГц представляється у вигляді 15 шарів із граничними частотами: (100÷400) Гц, (400÷800) Гц, (800÷1200) Гц, (1200÷1600) Гц, (1600÷2000) Гц, (2000÷2400) Гц, (2400÷2800) Гц, (2800÷3200) Гц,

(3200÷3600) Гц, (3600÷4000) Гц, (4000÷4400) Гц, (4400÷4800) Гц, (4800÷5200) Гц, (5200÷5600) Гц, (5600÷6000) Гц. Для кожного шару на сегменті в заданих частотних границях обчислюються параметри функцій Аньєзі, Гауса або ін., вирішуючи задачу апроксимації вихідного спектрально-часового представлення як суму локонів

$$Z(\omega_k, T_{IN}) = \sum_{i=1}^n G_i[W_{gij}(q_{ij}, \omega_k), W_{ij}(\alpha_{ij}, \beta_{ij}, c_{ij}, \omega_k), h_{ij}(\tau_{ij}, T_{ij}, d_{ij}, t_{IN})] \quad (4)$$

при $Z(\omega_k, T_{IN}) \rightarrow S(\omega_k, T_{IN})$.

За обчисленими значеннями параметрів локонів

$q_{(1,1)}, a_{(1,1)}, c_{(1,1)}, \beta_{(1,1)}, b_{(1,1)}, d_{(1,1)}, T_{(1,1)}$ будується п'ять груп описів

еталонних послідовностей кожного слова (складу):

- 1) максимуми функцій локона;
- 2) синтезоване спектрально-часове представлення;
- 3) бітове представлення із синтезованого спектрально-часового представлення;
- 4) двовимірний сплайн-опис спектрально-часового представлення на основі двовимірного фільтра [2];
- 5) вихідне спектрально-часове представлення $S(\omega_k, t_l)$.

При розпізнаванні реалізована ієрархія розпізнавання (3). Проста схема розпізнавання для словників до 1000 слів оперує з представленнями

Вид 2) → вид 5) ,

для яких задана внутрішня ієрархія, утворена послідовністю шарів у граничних частотах (100÷2000) Гц → (100÷3000) Гц → (100÷4000) Гц → (100÷4000) Гц → (100÷5000) Гц → (100÷6000) Гц.

На кожному кроці внутрішньої ієрархії реалізується розподіл обсягу розпізнаваної підмножини на дві.

Повна схема містить у собі всі параметричні представлення:

Вид 1) → вид 2) → вид 3) → вид 4) → вид 5).

Зіставлення за представленням «вид 1)» має найбільшу швидкодію, зіставлення по представленнях «вид 2)», «вид 5)» має найбільшу надійність. Кожен рівень зіставлення поставляє наступному половину еталонних послідовностей, звужуючи розпізнавану підмножину. На останньому етапі приймається остаточне рішення. Експериментальні дослідження на словнику у 500 слів виявили високу стійкість одержуваних правильних відповідей розпізнавання. У момент

навчання надійність розпізнавання становила не менш 98%. У процесі експлуатації на окремі слова виконується донавчання. Етап навчання полягає в накопиченні еталонних параметрів слів (слів-складів), вимовлених по одному разу, донавчання – додавання одного еталона ненадійно розпізнаваного слова.

Задача проектування бази знань аналізу і розпізнавання мовлення

Аналіз результатів у діючій системі розпізнавання виявив значну обмеженість алгоритмів і застосовуваних параметрів із причини жорсткості їхньої структури й неможливості підстроювання системи на кроці виконання з метою підвищення надійності розпізнавання та адаптації системи до ситуаційної варіативності мовлення. Подібну задачу можна і потрібно вирішувати в середовищі баз знань, керуючи системою розпізнавання на кроці виконання.

Задача побудови бази знань (БЗ) визначається цілями, що мають бути досягнуті в результаті її реалізації. Цілі створення БЗ – керування об'єктом та процесами, які в ньому відбуваються, що забезпечує досягнення деякого найкращого співвідношення між параметрами, які характеризують об'єкт і процеси, виявляють фактори, що викликають відхилення цих параметрів.

Об'єкт керування характеризується складом і призначенням як його частин, так і об'єкта в цілому, що визначає його статичний стан у вигляді інформаційного вектора $S(p_1, p_2, \dots, p_n)$ взаємозв'язку його параметрів p_l .

Процеси, що характеризують об'єкт, описуються функціями часу і параметрів динаміки об'єктів $f(t, p_1, p_2, \dots, p_n)$.

Задача керування може бути вирішена як задача моделювання передатної функції W об'єкта у виді сукупності його частин зі своїми передатними функціями W_i і як моделювання процесів у вигляді інформаційних потоків, що характеризують об'єкт у деякі фіксовані моменти часу $Y_{ki}(t_j, p_1, p_2, \dots, p_l, \dots, p_n)$, де k – номер частини об'єкта, i – номер інформаційного потоку в k -й частині об'єкта, p_l – l -й параметр як функція часу i -го інформаційного потоку в k -й частині об'єкта.

Задача практичного порівняння сигналів

Розглянемо задачу зіставлення двох сигналів на практиці. Порівняння будемо проводити на двох реалізаціях слова «чотири».

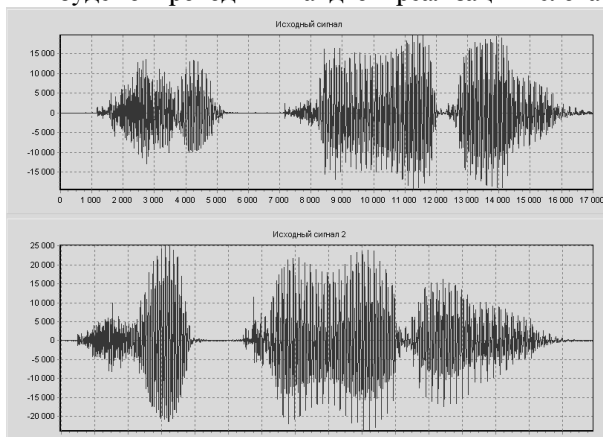


Рис. 1. Вид сигналу слова «чотири» у двох варіаціях

Після деяких стандартних перетворень із сигналом до отриманої матриці можна застосувати модифікований локон Аньєзі. Це перетворення у деякому сенсі подібне до перетворення Фур'є із тією лише різницею, що за базисну функцію обирається нетригонометрична функція.

Після цього можна подивитись на дві реалізації слова «чотири» у спектрально-часовому представленні.

Залежно від вимог до швидкості обчислення існують декілька підходів до отримання необхідних значень.

Перший із них полягає у пошуку максимального піку на кожній ітерації і відліку колоколу Аньєзі від цієї точки (де значення $S(\omega, t)$ набуває максимуму).

Інший підхід полягає в розбитті координатної площини на невеликі сегменти і підрахунку функції Аньєзі в кожному з них послідовно. Такий підхід є дещо швидший.

Наступним логічним кроком є порівняння отриманих величин. Спочатку розглянемо результат операції віднімання.

Як можна бачити з рис. 4, різниця є доволі невеликою і сигнали схожі. Іншою картина буде в разі порівняння діленням.

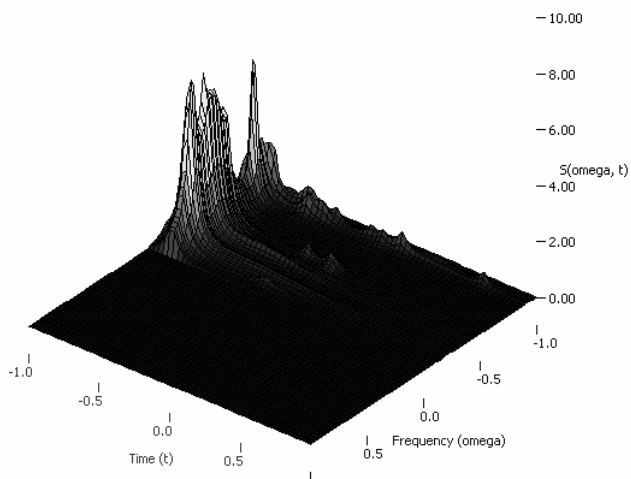


Рис. 2. Спектрально-часове представлення першої реалізації слова «чотири»

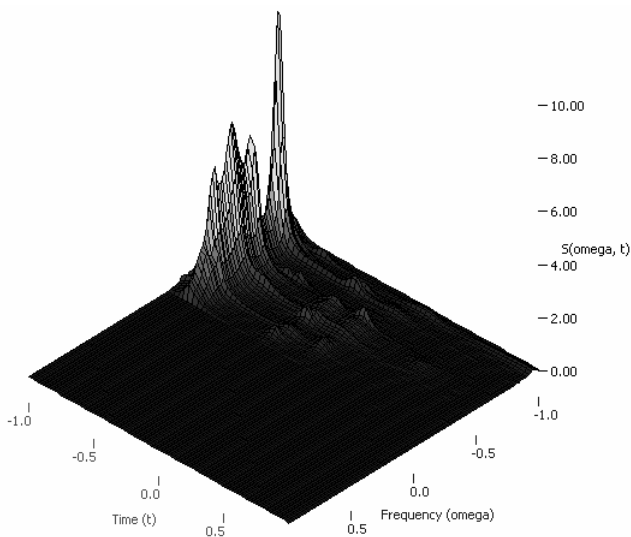


Рис. 3. Спектрально-часове представлення другої реалізації слова «чотири»

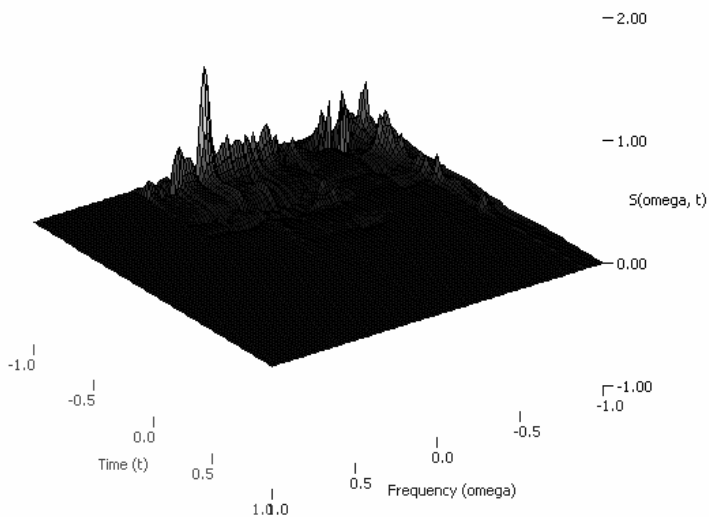


Рис. 4. Різниця спектрально-часового представлення (результат операції віднімання у збільшеному масштабі)

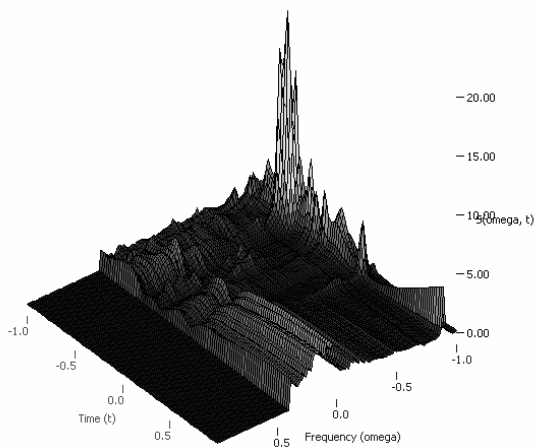


Рис. 5. Відношення спектрально-часових представлень (результат операції ділення)

Таким чином, ми отримали матрицю – передаточну функцію, що ідентифікує людину унікальним чином. Тобто виділили параметричне представлення для конкретної людини, що виражає певну емоцію.

Висновки. Відповідно до функції розпізнавання, технології проектування і загальної гіпотези про мовний сигнал як ієрархічну послідовність станів з обмеженими відхиленнями і необхідності його багаторівневої апроксимації як нелінійної функції побудована система виділення параметричного представлення мови людини.

Бібліографічні посилання

1. **Карпов О.М.** Вступ до курсу «Технологія проектування пристроїв розпізнавання промови»: навч. посіб. / О.М. Карпов. –Д.: РВВ ДДУ, 2000. – 64с.
2. **Карпов О.Н.** Технология построения устройств распознавания речи: монография / О.Н. Карпов. – Днепропетровск: Изд-во ДНУ, 2001. – 190 с.
3. **Карманов В.Г.** Математическое программирование / В.Г.Карманов. – М: Наука, 1975. – 272 с.
4. **Попов Э.В.** Экспертные системы. Решение неформализованных задач в диалоге с ЭВМ / Э.В.Попов. – М.: Наука, 1987. –288 с.
5. Компьютерные технологии распознавания речевых сигналов / [О.Н.Карпов, А.Г.Габович, Б.Г.Марченко, В.А.Хорошко и др.]. –К.: ООО «ПолиграфКонсалтинг », 2005, – 138 с.
6. **Douglas-Cowie** A New Emotion Database: Considerations, Sources and Scope / E.Douglas-Cowie, R.Cowie, M.Schroder. – Newcastle, 2000.

Надійшла до редколегії 12.10.2013 р.